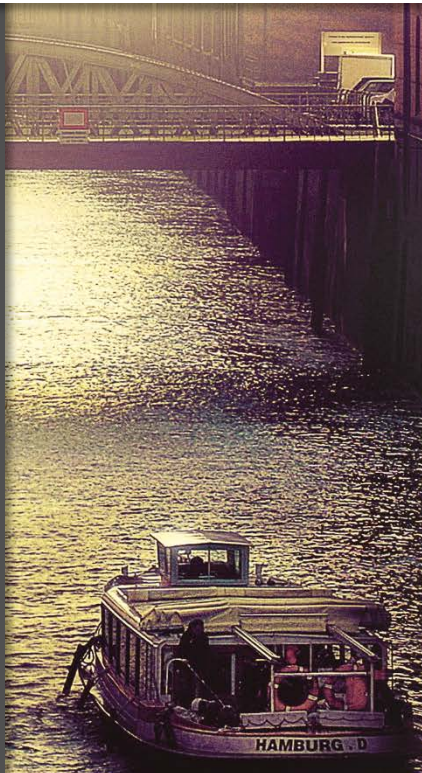


Benjamin Kormann, Susanne Altendorfer-Kaiser

Influence of Patterns and Data-Analytics on Production logistics



CC-BY-SA 4.0

Published in: Digitalization in Supply Chain Management and Logistics
Wolfgang Kersten, Thorsten Blecker and Christian M. Ringle (Eds.)
ISBN 9783745043280, Oktober 2017, epubli

Influence of Patterns and Data-Analytics on Production logistics

Benjamin Kormann¹, Susanne Altendorfer-Kaiser²

1 – Robert Bosch GmbH

2 – Montanuniversität Leoben

The flow of information is an essential part of Industry 4.0, as more reliable process data is available and subsequent hardware changes provide for processing power to enable large-scale data analysis. Due to the fact that most data analytics and big data frameworks presume that Data-Mining- and Data-Analytics-Activities are conducted in form of projects, this paper focuses on the integration of data analytics and data mining into operational processes and the resulting consequences of the organization. Therefore a framework to implement data analytics workflows in production logistics to improve decision-making and processes is presented. By integrating data analytics workflows in production logistics applying the presented framework, more resources can be devoted to proactively discover and counteract possible bottlenecks or constrictions instead of resorting to firefighting and taskforce-activities. The methodology to derive such framework consists of developing and implementing data analytics use-cases along the supply chain in production logistics according to current big-data and data-analytics frameworks in cooperation with a large automotive supplier and modifying current frameworks and approaches to fit the company's requirements.

Keywords: Knowledge Discovery in Databases; Data Analytics; Process model; Supply Chain Analytics

1 Introduction

Current initiatives and research approaches investigate intelligent, autonomous and decentralized subsystems that should lead to more competitive production and logistics processes - often revered to as "Industry 4.0" or "Internet of Things". These initiatives are not only driven by new technologies and methods of technical integration, in this context the information element is an essential asset for successful business: every business process - either service or production processes - requires a fundamental perception of information management to implement scientific logistics approaches to achieve an optimized supply chain (Jamil, Soares and Pessoa, 2017). Thus, Altendorfer-Kaiser (2015) states "Therefore the effective and economical integration of information and decision-making bodies is relevant".

Data and information are currently omnipresent and this oversupply of data and information for the business environment obtains more disadvantages than potentials for a company. In this context Information logistics becomes a cornerstone for companies: The goal of information logistics is to deliver the right information, in the right format, at the right place at the right time for the right people at the right price. (Uckelmann, 2012) Therefore an efficient information supply management is essential. However, information does not exist without data. And this is even important for the production industry as here data, information and generated insight based on the generated information are a major cornerstone for successful business in order to manage production processes and the resulting supply chains. (Fosso Wamba S., Akter S., 2015; Gimenez Isasi, Morosini Frazzon and Uriona, 2015; Militaru, Pollifroni and Ioanid, 2015; Meudt, et al., 2016; Corte-Real, Oliveira and Ruivo, 2017; Zhang, et al., 2017b; Zhang, et al., 2017a) Therefore this paper deals with the importance of big data in terms of production logistics and how big-data and analytics principles can be integrated in business processes extending existing knowledge discovery process models and which potential benefits can be derived by applying said model to business situations in production logistics.

The remainder is structured as follows: Section 2 provides background material on production logistics, big data analytics, smart data analytics, supply chain analytics, information and knowledge discovery process models. Section 3 discusses a possible extension of existing Knowledge discovery process models to ensure a lasting integration of big data analytics into business processes in production logistics. Section 4 discusses current application of the proposed model

in use-cases in the automotive supplier industry. Section 5 provides a summary and concluding remarks.

2 Background

2.1 Production logistics

Production logistics will be defined in this article as all operational purchasing, in-house material flow, in-house material handling, operational distribution and information flow processes which do need to be organized, controlled, executed and optimized in order to supply necessary raw-materials, perform manufacturing operations and physically distribute finished goods to customers. (Arnold, 2006; Günther and Tempelmeier, 2014; Wang, et al., 2016; Krieger W., 2017). Therefore, in order to facilitate further references to the mentioned process the SCOR-Model reference framework will be used to cluster the activities into Plan-, Source-, Make- and Deliver-Processes in Production Logistics (Bolstorff, Peter A., Rosenbaum, Robert G., Pohula, Rolf G., 2007; Supply-Chain Council, 2008) effectively defining production logistics as the management of in-house supply chains supplying manufacturing activities. In the context of Industry 4.0 many manufacturing related activities are being digitalized and products or manufacturing equipment are transformed in to cyber-physical systems which constantly gather data. (Schöning H., 2017) The resulting data can be considered Big Data, as the gathered data is being generated in near real-time. (Schöning H., 2017) In order to cope with such systems, pattern identification and other analytics tools related to Big Data serve as a viable way to handle complexity (Wehberg G., 2016) and in turn by actively handling complexity creating a basis which ensures future competitiveness and adaptability. (Schuh, Krumm and Amann, 2013; Leveling, Edelbrock and Otto, 2014)

2.2 Data and Information

In order to talk about information in general and for logistics in particular, it is important at this point to define the terms data and information to connect information and logistics in a more appropriate way. One approach can be seen in Figure 1.

Data: The noun data is defined as facts and statistics collected together for reference or analysis. The term itself comes from the Latin plural of “datum”.

Today organizations generate large amounts of multi-spectral data. In view of its discrete form, data in itself may not be very useful, so it is often referred to as the original knowledge asset. When data is processed into a context, it becomes information. (Bali, Wickramasinghe and Lehaney, 2009)

Information: For this paper the relevant definition of information is defined as something that is conveyed or represented by a particular arrangement or sequence. The term information origins in the Latin verb “informare” (in English “to inform”), which means ‘to give form’ or ‘to form an idea of’. Furthermore the Latin noun “informatio” had already had the meaning of concept and idea”. (Bali, Wickramasinghe and Lehaney, 2009) mentioned that information is data that has been arranged into a meaningful pattern and thus has an identifiable shape. An example is a report created from intelligent database queries. (Bali, Wickramasinghe and Lehaney, 2009) also found that information and communication technology not only increase the communication abilities with data but also accelerate the transferring and processing of this data into information.

When talking about data and information it is also interesting to have look on IBMs “Business Information Maturity Model”, which defines five levels of data management and is shown in figure 2.

The focus of data at the lowest level is from an operational perspective. At the next level, the Information is used to manage the company. The Information becomes a strategic asset at the next level. At the next level the Information becomes a form of special expertise. Finally, at the top level, the Information is what give the company a competitive advantage and therefore often needed to be protected against external actors. (Arlbjørn and Haug, 2010)

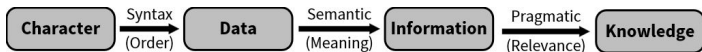


Figure 1: Interconnection between data, information and knowledge (Auer, 2008)

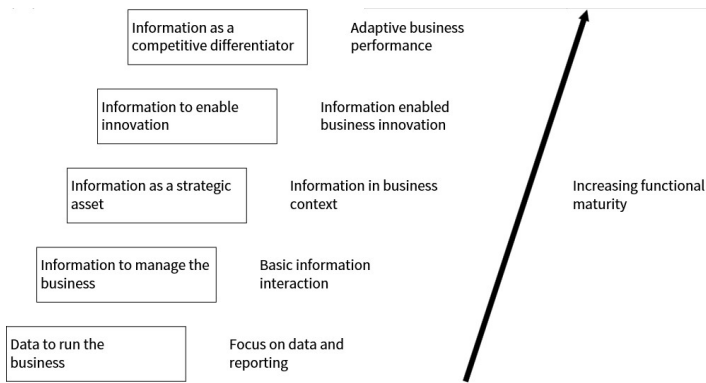


Figure 2: Business Information Maturity Model (Arlbjørn and Haug, 2010)

2.3 Big and Smart Data Analytics

(Provost and Fawcett, 2013) define “Big Data as ‘datasets that are too large for traditional data-processing systems and that therefore require new technologies’”. Therefore big data is mainly produced by machines and thus often represented as machine data, too. According to (Cooper M., 2012) ‘Big data is where the data volume, acquisition velocity, or data representation limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant horizontal scaling for efficient processing.’ Such views focus on the data domain and are therefore insufficient to cover all topics faced in managing Production logistics processes, as for example complexity or analyzing the data gathered is not covered in the definitions above. In order to connect the different characteristics of Big Data to Production logistics the mentioned view needs to be expanded further creating an integrated view of the topic, which includes dependencies of characteristics, business intelligence, statistics, data and characteristics clustering (Wu, Buyya and Ramamohanarao, 2016; Lee, 2017). In doing so the characteristics of Big Data can be clustered in three domains. (Wu, Buyya and Ramamohanarao, 2016)

Data Domain

The Data Domain covers the Variety, Velocity and Volume of Big Data. (Wu, Buyya and Ramamohanarao, 2016) Volume measures the amount of data an organization generates and has already generated, Velocity measures the speed of data generation and Variety measures the different types of data accessible to potential analysis. (Kaisler, et al., 2013) These characteristics cover the descriptions of Provost and Fawcett as well as Cooper's and Mel's view of the subject.

Statistics Domain

The Statistics Domain covers Veracity, Validity and Variability. The characteristics cover issues in data quality, reliability, complexity, variation, decay and data uncertainty. (Wu, Buyya and Ramamohanarao, 2016)

Business Intelligence Domain

The Business Intelligence Domain includes the topics of Visibility, Verdict and Value. (Wu, Buyya and Ramamohanarao, 2016) By adding such characteristics views on decision making, value of data for the business namely extracting valuable information and data hindsight, insight and foresight are also covered. (Wu, Buyya and Ramamohanarao, 2016)

Integrating all characteristics implies that they cannot be viewed independently from each other. (Lee, 2017) Increases in volume, variety and velocity increase complexity, variability and value but decrease veracity. (Lee, 2017) In order to develop a process model, all aspects need to be considered accordingly as the goal is to ensure a lasting integration of Big Data into business processes in production logistics.

Additionally, the topic of value must be further investigated. Big Data itself does not necessarily generate value for the company. The generated information must be presented, aligned with current processes, kept safe, enhanced with previous experiences and aggregated accordingly in order to enable human interactors to decide based on results efficiently and effectively. (Reich R., Mohanty S., Litzel N.; Kaisler, et al., 2013; Coffey L., 2014; Heuring W., 2015; Jähnichen S., 2015; Diesner M., 2016) Big Data Analytics becomes Smart Data Analytics. Smart Data Analytics thus requires also to consider information management aspects which deepen

the information process view perspective to be able to manage the information and the consequent distribution of such generated insights by data analysis processes. Additionally, the connection with smart data analytics and supply chain management has to be clarified in order to elaborate potential benefits of data mining and data analytics in supply chains.

2.4 Supply Chain Analytics

Applying smart data in supply chains or production logistics domain results in supply chain analytics which are defined by applying big data and data analytics in the context of logistics and supply chains. (Wang, et al., 2016) Additionally Supply Chain Analytics can be seen as dynamic capabilities of a company. (Chae and Olson, 2013) These capabilities are analytical, IT-enabled, improve supply chain performance. (Chae and Olson, 2013) They consist of a data management, an analytical supply chain process and a supply performance management capability. (Chae and Olson, 2013) These capabilities are necessary to gain useful information and knowledge from supply chain data. (Chae and Olson, 2013) The use of the gathered information and knowledge results in optimized operational, tactical and strategic decisions in the plan-, source-, make-, deliver- and return-phases of supply chains. (Souza, 2014) The necessary analytics-capabilities can be divided into four sections: generating information, insights, decisions and actions (Sivarajah, et al., 2017) These sections are (Sivarajah, et al., 2017):

Descriptive analytics Descriptive analytics provide a basis to understanding what has happened in supply chains.

Inquisitive/diagnostic analytics Diagnostic analytics generate insight into why the results derived in the descriptive phase have occurred.

Predictive analytics Predictive analytics is based on the complete understanding of past events generated in the descriptive and diagnostic phase and based on these findings anticipates future events within certain statistical boundaries.

Prescriptive analytics Prescriptive analytics provides a basis to decision-making processes helping decision-makers gain objective and transparent views on historical events in combination with results from the predictive phase. Questions like "What now?" can be answered

Pre-emptive analytics This step recommends based on the decisions made possible courses of action to prevent certain events derived in the predictive phase.

The application and implementation of supply chain analytics must be measured to derive measures for each stage to take in order to apply supply chain analytics most efficiently and effectively to improve business processes. (Arunachalam, Kumar and Kawalek, 2017) The proposed methodology by Arunachalam, Kumar and Kawalek (2017) for measuring divides the capabilities of supply chain analytics in analytics, visualization, data-driven culture, data generation and data integration. Measuring all these capabilities the following stages are identified in Figure 3. (Arunachalam, Kumar and Kawalek, 2017)

Companies must therefore initiate supply chain analytics with data gathering and being able to understand past events of internal supply chains with data. (Initiation stage) This must then be expanded to predictions inside internal supply chains and the inclusion of external data. (Adoption stage) The routinization stage consists of generating knowledge and value of the derived information effectively turning data into a competitive advantage. The stages in figure 3 show that data and analytics capabilities go hand-in-hand in order to derive information, knowledge and value from data.

2.5 Information Management Models

Information Management Models aim to provide reference activities to define one possible efficient and effective way of managing the resource information in organizations. (Krcmar, 2015) There exist some well described Information Management Models, whereas everyone has its individual focus. Three different approaches are (Stahlknecht and Hasenkamp, 2002), Heinrich L. J., Stelzer D. (2011) and Mertens, et al. (2012). Stahlknecht and Hasenkamp (2002) focus on Information as the third production factor and concentrate on the technical information acquisition and storage. Mertens, et al. (2012) also address the technical focus. Heinrich L. J., Stelzer D. (2011), however, focus on the management aspects of information management and aims to establish an information management infrastructure that supports decision-making. In the context of this paper only the model of Krcmar (2015) is analysed in detail as it covers all characteristics mentioned with smart data. Krcmar's model divides information management in four sections. These are executive activities, information economics (demand,

supply and consumption), software management and hardware management. (Krcmar, 2015) These sections respectively include: (Krcmar, 2015)

Executive activities Such activities include strategic aspects of information management, IT-Governance, information processes, human resources, financial controlling and IT-Security.

Information economics Information economics activities result in exact knowledge of information demand, information supply and by consequently combining the two deducting the necessary information use of the organization.

Software and hardware management These activities ensure that life-cycles of applications and hardware are generated, maintained and observed. Additionally these sections ensure that the data necessary to satisfy demand is supplied in the most efficient and effective manner possible.

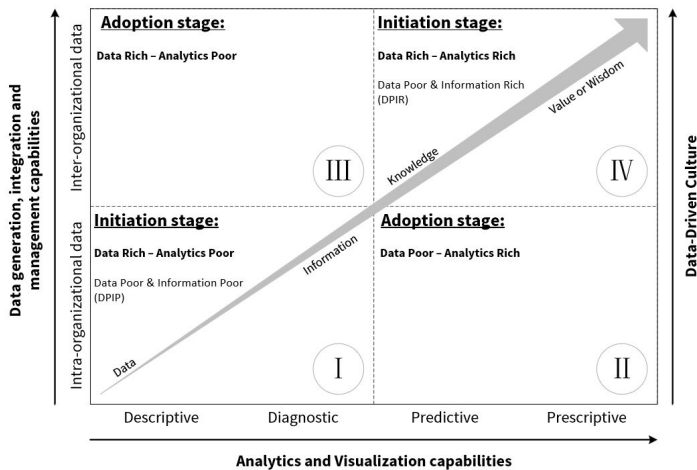


Figure 3: Supply Chain Analytics capability framework (Arunachalam, Kumar and Kawalek, 2017)

The presented information model does not include a section relevant to the creation of value for the organization based on the information provided. Therefore the current landscape on big data reference process models must also be analyzed in depth.

2.6 Big Data and Knowledge discovery process reference models

In order to work with Big Data and master its characteristics certain process steps need to be considered in order to create knowledge about one specific domain. Furthermore the analytics perspective needs to be incorporated into Big Data in order to consider that identified patterns are transformed in usable process knowledge. Such processes can be summarized by the concept of Knowledge Discovery in Databases. (Fayyad, U., Piatetsky-Shapiro G., Smyth G. P., 1996) The academic world and industries alike have developed different kinds of approaches on how and what to face when trying to gain knowledge out of data. The concepts analyzed in the context of this paper are:

- Knowledge Discovery in Databases (KDD) Process
(Fayyad, U., Piatetsk-Shapiro G., Smyth G. P., 1996).
- Information Flow in a Data Mining Life Cycle
(Ganesh, et al., 1996; Kopanakis I., 1999).
- SEMMA
(SAS Institute Inc, 1997)
- Refined KDD paradigm
(Collier, K., Carey, B., Grusy, E., Marjaniemi, C., Sautter, D., 1998).
- Knowledge Discovery Life Cycle (KDLC) Model
(Lee and Kerschberg, 1998).
- Cross-Industry-Standard Process for Data Mining (CRISP-DM)
(CRISP-DM Consortium, 2000).
- RAMSYS
(Moyle; Blockeel and Moyle, 2002).
- Generic Data Mining Life Cycle (DMLC)
(Hofmann, 2003; Hofmann and Tierney, 2009).

- Ontology Driven Knowledge Discovery Process (ODKD) (Gottgroy, 2007).

- Adaptive Software Development-Data Mining (ASD-DM) Process Model (Alnoukari, Alzoabi and Hanna, 2008).

- A Data Mining & Knowledge Discovery Process Model (Marbán Ó, Mariscal G., Segovia J., 2009).

- AgileKDD (Nascimento G., 2012).

- Big Data Analytics Activity Reference Framework (Yew, 2015).

- Knowledge Discovery in Data Science (Grady, 2016).

All mentioned methodologies focus on the data mining tasks and only partially consider how to transfer the knowledge gathered into the organization. Furthermore, all methodologies presume that all process steps are performed as projects and not continuous improvement tasks to optimize business processes constantly. Additionally, no explicit loops are considered to stop the cycle once it becomes evident that with the current quality and availability of prerequisites such as data itself for example the project needs to be modified and transformed into a change process. This process then aims to transform the involved business processes in such a manner that they fulfill the prerequisites and provide the basis to improve business processes with data mining and knowledge discovery methods. Pivk, A., Vasilecas O., Kalibatiene D., Rupnik R.(2013) propose a methodology to use data mining to optimize business processes. Adapting processes to incorporate gathered knowledge is being considered, yet this model does propose to perform loops when required, certain prerequisites are not met or model deployment uncovers possible changes necessary to business processes in order to perform optimizations with data mining and knowledge discovery in databases.

3 Extension of existing approaches to data mining and knowledge discovery in databases

In order to solve the weaknesses mentioned above, a combination of the AgileKDD model, the Generic Data Mining Life Cycle, the Data Mining & Knowledge Discovery Process Model by Marban (2009), the proposed implementation of datamining to perform business process optimization (Pivk, A., Vasilecas O., Kalibatiene D., Rupnik R., 2013), the Knowledge Discovery in Data Science approach (Grady, 2016), the Big Data Analytics Activity Reference Framework (Yew, 2015) and Krcmar's Information Management Model (Krcmar, 2015) is proposed. Combining these models, the following reference activities must be incorporated when trying to derive a process reference model suitable to integrate knowledge discovery results into business processes under the aspect of performing loops when prerequisites of stages are not met:

- Identify potentials in business processes where data mining activities can improve performance. (Identify potentials)
- Provide a transparent and decisive process to evaluate benefits and necessary resources for chosen activities ideally expressed in monetary units. (Benefits)
- Apply traditional data mining activities which potentially can generate necessary knowledge to realize identified performance improvements based on a predefined selection matrix customized to production logistics needs. (Data Mining)
- Apply chosen techniques within a standardized environment mainly consisting of a big-data-platform which provides all necessary hardware, software, a knowledge repository documenting previous activities, a model repository documenting previous modelling approaches and access to various data sources in a timely manner. (Deployment)
- Continuously evaluate the progress of activities related to all necessary steps to allow modifications to previous steps once disrupting events are uncovered during the execution of tasks. (Evaluation)
- Provide suitable organizational methods and means to preserve the generated knowledge, share it across the organization and ensure experi-

ences made in the application of the results are considered in future iterations of the process. (Integration)

These steps are only to be executed with an implemented and fully operational big-data platform allowing the use of standardized hardware, software, scaling, interfaces, data sources, knowledge repositories, model repositories, data governance protocols, privacy protection and security measures.

The "identify potentials"-phase consists of a thorough analysis of the business processes to be improved with a focus on the information flows and the consequent tasks mainly triggered by generated information during the execution of the processes. Such focus is necessary as the model aims to optimize information flows and due to the focus on information all data understanding topics are also covered. This phase can be compared to Hofmann's phase of constantly looping through the business understanding and data preparation phase until the understanding of both domains is sufficient to set the objectives and hypothesis clearly to proceed with data processing. (Hofmann, 2003) Furthermore as the objective is to optimize the process a change process must be conceived once the desired results are generated. This aspect also incorporates the Assessment- and Business-Process-Renovation-phase in the proposed approach of (Pivk, A., Vasilecas O., Kalibatiene D., Rupnik R., 2013) excluding application analysis, design, development and design. Additionally, the first six steps of the Big Data Analytics Activity Reference Framework by Yew (2017) are considered.

The benefits-phase determines the necessary resources to optimize the chosen business processes with supply chain data analytics and weigh them against potential benefits gained from the generated insight of the knowledge discovery process. This step mainly incorporates Marban's (2009) approach to data mining where software engineering and project management practices are incorporated. (Marbán Ó, Mariscal G., Segovia J., 2009) If the benefits-process is triggered by the processes downstream the necessary changes to be made must be assessed and weighted against the resources necessary to implement them.

The Data Mining phase consists of the core Knowledge Discovery process covering data transformation, pre-processing, choosing suitable data mining techniques, developing a model, apply the model and evaluate results. (CRISP-DM Consortium, 2000; (Marbán, et al., 2007)

The deployment-phase consists of adequately scaling the results and the model to suit the business process to be optimized within the limits of the platform.

Furthermore, the generated model and the gathered knowledge are to be transferred to the respective repositories on the platform resembling the steps in the Hofmann (2003) approach.

The evaluation- phase covers the maintenance and monitoring of the deployed solution and should verify the benefits derived in the identify-potentials-phase. The focus should be put on the improvement of decision-making based on the generated knowledge. It can also be considered as a pilot-implementation-phase.

The last step consists of integrating the improved process into the organization by developing a roll-out-plan which consists of activities in the Integral process domain cited in Marbán's model. (Marbán Ó, Mariscal G., Segovia J., 2009) These are, for example, training of employees or create a documentation.

The loops in shown in Figure 4 allow for the mentioned weakness of transforming the process into a change process to ensure the readiness of the process to incorporate data mining as an improvement tool. This change process is then started during the benefits phase if the potential benefit is sufficient to justify the necessary changes in business processes. The change process can be triggered in various other phases of the model.

The model therefore enables companies by repeatedly executing the steps to build up their supply chain analytics capabilities and steadily implement truly data-driven supply chains as it can be applied in the four stages of Supply Chain analytics and because of the loops dynamically adapted to suit the needs in every phase.

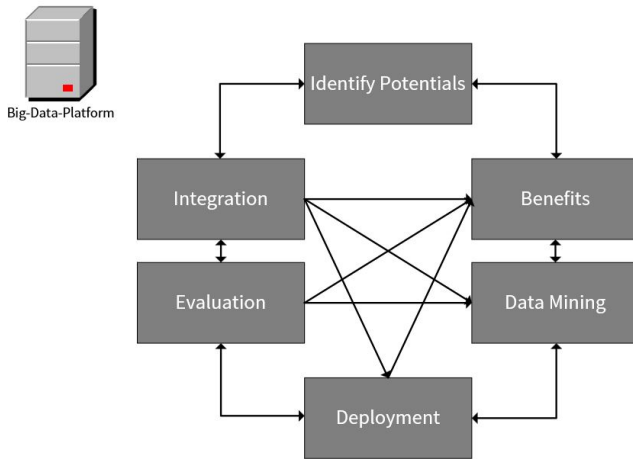


Figure 4: Knowledge Discovery Model for Production Logistics

4 Application in the automotive supplier industry

The developed model was applied in cooperation with a large international automotive supplier with the goal to integrate data mining and knowledge discovery in databases within its production logistics of the electronics manufacturing domain.

The framework was developed while trying to integrate data analytics in business process in the plan-, source-, make- and delivery-domains with current methods mentioned in Section 3. During the application of the methods the mentioned weaknesses became evident resulting in the development of the new reference model. The application of the model to the use-cases is still ongoing.

Currently the focus is to thoroughly define, implement and the "Identify potentials"- and "Benefits"-phases of the model. The "Identify potentials"-phase in the application domain consists of a combination of three approaches to systematically determine possible applications of Supply Chain Analytics. These are:

1. Interviews

The first and currently most used approach is to interview managers about current issues regarding information deficits and labor-intensive processes necessary to provide the organization with decision-relevant information. The results are then evaluated and processed in order to determine if theoretical benefits can be generated by applying descriptive analytics, diagnostic-, predictive-, prescriptive- or pre-emptive analytics. The general rule currently being applied is that possible use-cases only generating descriptive insight are not being pursued as this kind of activity is mostly connected to reporting-topics which require a completely different kind of methodological approach. If theoretical benefits are found, the next process step can be started.

2. Process analysis

As a prerequisite possible processes suitable for improvement must be chosen. This requires a rather superficial and fast way to analyze processes. In the automotive context the SIPOC-approach was chosen. The approach consists of describing the supplier, input, process, output and customer in a very short and precise way. (Toutenburg and Knöfel, 2009).

Based on the rough process analysis inefficiencies and optimization potentials are then identified. These inefficiencies and potentials are then evaluated by the possible benefits of the application of data analytics. The focus is to be laid on the input, output, suppliers and customers as the knowledge generated by the application of data analytics can either help to improve the information content of input/output or help to better understand the behavior of the customer/supplier. If the SIPOC-analysis has generated promising potentials a more detailed analysis of the process to be improved can be started.

This step requires a precise documentation of processes ideally using the BPMN-notation, as this notation also includes data flows. These flows provide basis for an analysis based on information management principles and by determining deviations of the current state to the target picture. The deviations are then further investigated in order to determine if the application of analytics-tools provides an improvement leading to a decrease of the determined deviation. If the reduction is deemed sufficient enough the next process step of the model can be started.

3. Benchmarking

Benchmarking activities in this context include extensive literature research and the compilation of locally implemented data-analytics solutions within the manufacturing network of the company. These two views must then be combined, generating a view on the subject containing internal and external activities. This view must then be used to determine possible use-cases using data analytics in a workshop with operational management.

Once possible use-cases are determined a benefit analysis is conducted in order to determine the economic benefits. This benefit analysis contains qualitative and quantitative aspects.

The qualitative aspects include compliance improvements, conformance improvements, decision-making basis improvements, increase in transparency, analytics maturity and risks regarding goal-achievements. The analytics maturity is evaluated using the desired data analytics level (descriptive-, diagnostic-, predictive-, prescriptive-, pre-emptive analytics). The further the use-case advances the higher the evaluation.

The quantitative aspects include rough estimates regarding costs of development, implementation and maintenance. The costs are then compared

against possible cost savings and performance improvements. Furthermore the qualitative aspects are included into the cost-saving-calculation with a factor.

No conclusive results can be presented at this moment in time, yet the approach of aligning data analytics activities with the available resources of the platform and implemented business processes has rendered great acceptance of the proceeding within the affected parts of the organization and first pilot-phases have shown great potential to transform production logistics into a data-driven endeavor seamlessly integrateable in other Industry-4.0-activities within the company. Furthermore, the operational, tactical and strategic decisions are already partially based on the results of the pilots which are now based on knowledge previously unused resulting in focusing available resources more efficiently and effectively on current business challenges. These changes in resource allocation are now used to proactively uncover potential disruptions to processes avoiding fire-fighting and task-forces activities.

5 Conclusion

The presented supply chain analytics and knowledge discovery model does by design align with current business processes and therefore can seamlessly be integrated into current continuous improvement activities which facilitates the acceptance of data analytics solutions among all parties involved. No definitive results can be presented at this time to prove the effectiveness and efficiency of the model in the context of production logistics. Current results of the running pilots have the desired managerial implications of being able to allocate resources to proactively discover and counteract potential process disruptions which can be discovered by data-based decisions enabled by the application of the presented knowledge discovery model. Information can then be used as a competitive differentiator. Furthermore the first two steps of the model have been systematically applied with success in order to determine possible pilot use-cases to test and develop the model further.

The model does have its limitations as the application has been limited to in-house supply-chain-processes in the automotive supplier industry. The implementation of the model does proceed accordingly, yet only the first two steps in the model have been thoroughly defined and applied to real-world applications. The other

steps are currently highly theoretical and in pilot phases. Further research has to be performed to specify the remaining processes in the model more thoroughly in order to make them applicable to other domains in supply chain management. Additionally, the platform needs to be specified further.

References

- Alnoukari, M., Z. Alzoabi, and S. Hanna (2008). "Applying adaptive software development (ASD) agile modeling on predictive data mining applications: ASD-DM methodology". In: *2008 International Symposium on Information Technology*. IEEE, pp. 1–6.
- Altendorfer-Kaiser, S. (2015). "Information Logistics Means to Support a Flexible Production?" In: *Advances in Production Management Systems: Innovative Production Management Towards Sustainable Growth*. Ed. by S. Umeda, M. Nakano, H. Mizuyama, N. Hibino, D. Kiritsis, and G. von Cieminski. Vol. 459. IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing, pp. 414–421.
- Arlbjörn, J. S. and A. Haug (2010). *Business process optimization*. 1. edition. Århus: Academica.
- Arnold, D. (2006). *Intralogistik*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Arunachalam, D., N. Kumar, and J. P. Kawalek (2017). "Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice". In: *Transportation Research Part E: Logistics and Transportation Review*.
- Auer, T. (2008). "Wissensmanagement: Reizwort oder zeitgemäße Notwendigkeit?" In: *Controller Leitfaden 12*.
- Bali, R. K., N. Wickramasinghe, and B. Lehane (2009). *Knowledge management primer*. Routledge series in information systems. New York and London: Routledge.
- Blockeel, H. and S. Moyle (2002). "Collaborative data mining needs centralised model evaluation". In: *Proceedings of the ICML-2002 Workshop on Data Mining Lessons Learned*, pp. 21–28.
- Bolstorff, Peter A., Rosenbaum, Robert G., Pohula, Rolf G. (2007). *Spitzenleistungen im Supply Chain Management*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Chae, B. and D. L. Olson (2013). "Business analytics for supply chain: a dynamic-capabilities framework". In: *International Journal of Information Technology & Decision Making* 12.01, pp. 9–26.
- Coffey L. (2014). *Smart Data - What is it & How is it Different From Big Data?*
- Collier, K., Carey, B., Grusy, E., Marjanemi, C., Sautter, D. (1998). *A perspective on Data Mining*. Ed. by Center for Data Insight. USA.
- Cooper M., M. P. (2012). *Tackling Big Data*. Ed. by Federal Computer Security Managers' Forum. USA.
- Corte-Real, N., T. Oliveira, and P. Ruivo (2017). "Assessing business value of Big Data Analytics in European firms". In: *Journal of Business Research* 70, pp. 379–390.
- CRISP-DM Consortium (2000). *CRISP-DM 1.0: Step-by-step data mining guide*.
- Diesner M. (2016). "Massendaten im Fertigungsumfeld". In: *visIT[Big Data] - Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB*, pp. 6–7.
- Fayyad, U., Piatetsky-Shapiro G., Smyth G. P. (1996). "From data mining to knowledge discovery in databases". In: *AI Magazine*, pp. 37–54.

- Fosso Wamba S., Akter S. (2015). "Big data analytics for supply chain management: A literature review and re-search agenda". In: *Enterprise and Organizational Modeling and Simulation*, pp. 61–72.
- Ganesh, M., E.-H. Han, V. Kumar, S. Shekhar, and J. Srivastava (1996). *Visual Data Mining: Framework and Algorithm Development*.
- Gimenez Isasi, N. K., E. Morosini Frazzon, and M. Uriona (2015). "Big Data and Business Analytics in the Supply Chain: A Review of the Literature". In: *IEEE Latin America Transactions* 13.10, pp. 3382–3391.
- Gottgroy, P. C. M. (2007). "Ontology Driven Knowledge Discovery Process: A proposal to integrate Ontology Engineering and KDD". In: *PACIS*.
- Grady, N. W. (2016). "KDD meets Big Data". In: *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 1603–1608.
- Günther, H.-O. and H. Tempelmeier (2014). *Supply Chain und Operations Management*. 11., verbesserte Auflage. Vol. / Horst Tempelmeier; Hans-Otto Günther ; [Hauptbd.] Produktion und Logistik. Norderstedt and Norderstedt: Books on Demand.
- Heinrich L. J., Stelzer D. (2011). *Informationsmanagement*. Oldenburg: Oldenburg Wissenschaftsverlag.
- Heuring W. (2015). *Warum Big Data zu Smart Data werden muss!*
- Hofmann, M. (2003). "The development of a generic data mining life cycle (DMLC)". Ph. D. Ireland: School of Computing.
- Hofmann, M. and B. Tierney (2009). "An enhanced data mining life cycle". In: *2009 IEEE Symposium on Computational Intelligence and Data Mining*. IEEE, pp. 109–117.
- IARIA, ed. (2012). *The Seventh International Conference on Software Engineering Advances*.
- Jähnichen S. (2015). "Von Big Data zu Smart Data – Herausforderungen für die Wirtschaft". In: *Newsletter Smart Data August 2015*, pp. 1–2.
- Jamil, G. L., A. L. Soares, and C. R. M. Pessoa (2017). *Handbook of Research on Information Management for Effective Logistics and Supply Chains*. IGI Global.
- Kaisler, S., F. Armour, J. A. Espinosa, and W. Money (2013). "Big Data: Issues and Challenges Moving Forward". In: *2013 46th Hawaii International Conference on System Sciences*. IEEE, pp. 995–1004.
- Kopanakis I., T. B. (1999). "Visual data mining & modeling techniques". In: *University of Manchester Institute of Science and Technology - Department of Computation*.
- Krcmar, H. (2015). *Informationsmanagement*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Krieger W. (2017). *Definition Produktionslogistik*. Ed. by Gabler Wirtschaftslexikon.
- Lee, I. (2017). "Big data: Dimensions, evolution, impacts, and challenges". In: *Business Horizons* 60.3, pp. 293–303.
- Lee, S. W. and L. Kerschberg (1998). "A methodology and life cycle model for data mining and knowledge discovery in precision agriculture". In: *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218)*. IEEE, pp. 2882–2887.
- Leveling, J., M. Edelbrock, and B. Otto (2014). "Big data analytics for supply chain management". In: *2014 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, pp. 918–922.
- Marbán Ó, Mariscal G., Segovia J. (2009). "A Data Mining & Knowledge Discovery Process Model". In: *Data Mining and Knowledge Discovery in Real Life Applications*. Ed. by K. A. Ponce J. I-Tech Education and Publishing, Ch. 1.

- Marbán, Ó., G. Mariscal, E. Menasalvas, and J. Segovia (2007). "An Engineering Approach to Data Mining Projects". In: *Intelligent Data Engineering and Automated Learning - IDEAL 2007*. Ed. by H. Yin, P. Tino, E. Corchado, W. Byrne, and X. Yao. Vol. 4881. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 578–588.
- Mertens, P., F. Bodendorf, W. König, A. Picot, M. Schumann, and T. Hess (2012). *Grundzüge der Wirtschaftsinformatik*. 11. Aufl. 2012. Springer-Lehrbuch. Berlin, Heidelberg: Springer Berlin Heidelberg and Imprint: Springer.
- Meudt, T., M. P. Rößler, J. Böllhoff, and J. Metternich (2016). "Wertstromanalyse 4.0". In: *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb* 111.6, pp. 319–323.
- Militaru, G., M. Pollifroni, and A. Ioanid (2015). "BIG DATA IN SUPPLY CHAIN MANAGEMENT: AN EXPLORATORY STUDY". In: *Network Intelligence Studies* 6, pp. 103–108.
- Moyle, S. (n.d.). "RAMSYS - A methodology for supporting rapid remote collaborative data mining projects". In:
- Nascimento G., O. A. (2012). "AgileKDD: An Agile Business Intelligence Process Model". In: *The Seventh International Conference on Software Engineering Advances*. Ed. by IARIA, pp. 598–604.
- Pivk, A., Vasilecas O., Kalibatiene D., Rupnik R. (2013). *On approach for the implementation of data mining to business process optimisation in commercial companies*. The Free Library.
- Ponce J., K. A., ed. (2009). *Data Mining and Knowledge Discovery in Real Life Applications*. I-Tech Education and Publishing.
- Provost, F. and T. Fawcett (2013). *Data science for Business: What you need to know about data mining and data-analytic thinking*. First edition. Sebastopol, Calif.: O'Reilly.
- Reich R., Mohanty S., Litzel N. (n.d.). *Nach Big Data kommt Smart Data*.
- SAS Institute Inc (n.d.). *SEMMA: Data mining and the case for sampling*. SAS Institute Inc.
- Schöning H., D. M. (2017). "Big Smart Data – Intelligent Operations, Analysis und Process Alignment". In: *Handbuch Industrie 4.0 Bd.2*. Ed. by B. Vogel-Heuser, T. Bauernhansl, and M. ten Hompel. Springer Reference Technik. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 457–469.
- Schuh, G., S. Krumm, and W. Amann (2013). *Chefsache Komplexität*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Sivarajah, U., M. M. Kamal, Z. Irani, and V. Weerakkody (2017). "Critical analysis of Big Data challenges and analytical methods". In: *Journal of Business Research* 70, pp. 263–286.
- Souza, G. C. (2014). "Supply chain analytics". In: *Business Horizons* 57.5, pp. 595–605.
- Stahlknecht, P. and U. Hasenkamp (2002). *Einführung in die Wirtschaftsinformatik*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Supply-Chain Council (2008). *Supply-Chain Operations Reference-Model Version 8.0*. Pittsburg, USA.
- Toutenburg, H. and P. Knöfel (2009). *Six Sigma: Methoden und Statistik für die Praxis*. 2., verbesserte und erweiterte Aufl. Berlin and Heidelberg: Springer.
- Uckelmann, D. (2012). *Quantifying the Value of RFID and the EPCglobal Architecture Framework in Logistics*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Umeda, S., M. Nakano, H. Mizuyama, N. Hibino, D. Kiritsis, and G. von Cieminski, eds. (2015). *Advances in Production Management Systems: Innovative Production Management Towards Sustainable Growth*. IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing.
- Vogel-Heuser, B., T. Bauernhansl, and M. ten Hompel, eds. (2016). *Handbuch Industrie 4.0 Bd.3: Logistik*. Wiesbaden: Springer Science and Business Media and Springer Vieweg.

- Vogel-Heuser, B., T. Bauernhansl, and M. ten Hompel, eds. (2017). *Handbuch Industrie 4.0 Bd.2: Automatisierung*. Springer Reference Technik. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Wang, G., A. Gunasekaran, E. W. Ngai, and T. Papadopoulos (2016). "Big data analytics in logistics and supply chain management: Certain investigations for research and applications". In: *International Journal of Production Economics* 176, pp. 98–110.
- Wehberg G. (2016). "Big Data – Mustererkennung als Erfolgsfaktor der Logistik 4.0". In: *Handbuch Industrie 4.0 Bd.3*. Ed. by B. Vogel-Heuser, T. Bauernhansl, and M. ten Hompel. Wiesbaden: Springer Science and Business Media and Springer Vieweg, pp. 377–392.
- Wu, C., R. Buyya, and K. Ramamohanarao (2016). "Big Data Analytics = Machine Learning + Cloud Computing". In: *CoRR* abs/1601.03115.
- Yew, S. L. B. (2015). "Building use Cases With Activity Reference Framework for Big Data Analytics". In: *2015 9th International Conference on IT in Asia (CITA)*. IEEE, pp. 1–7.
- Zhang, Y., S. Ren, Y. Liu, T. Sakao, and D. Huisingh (2017). "A framework for Big Data driven product lifecycle management". In: *Journal of Cleaner Production* 159, pp. 229–240.
- Zhang, Y., S. Ren, Y. Liu, and S. Si (2017). "A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products". In: *Journal of Cleaner Production* 142, pp. 626–641.